Time dependent covariates in a competing risks setting

G. Diao A. N. Vidyashankar S. Zohar
S. Katsahian*

*Unité de Biostatistique et Recherche Clinique
Hôpital Henri-Mondor, Université Paris 12
Créteil, France

32th Annual Conference of the International Society for Clinical Biostatistics, 2011

Outline

Competing risk models

- Multiple causes for risk
- Can be characterized by the cause-specific hazard

$$\lambda_j(t|\mathbf{X}) = \lambda_j(t)\exp(\beta_j^T \mathbf{X}),$$

  where β is a set of regression coefficients and $\lambda_j(t)$ is the baseline hazard for the jth cause
- In the above, it is assumed that
  - hazard ratios are constant over time
  - covariates are time-independent or external time-dependent

A Motivating Example

European Bone-Marrow Transplantation (EBMT) Study

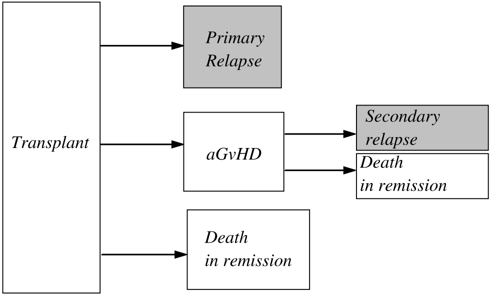
- The study included 541 patients
  - receiving allogeneic Unrelated Bone Marrow Transplant
  - less than 16 years old at time of transplant
  - had Acute Leukemia
- One objective of the study is to evaluate covariate effects on relapse accounting for competing causes of death
- Both internal time-dependent covariates (e.g., occurrence of aGvHD) and external time-dependent covariates (e.g., age) are present

- *External* time-dependent covariates versus *internal* time-dependent covariates:
  - The path of an external time-dependent covariate is generated externally. For example, age, levels of air pollution, etc.
  - The change of an internal time-dependent covariate over time is related to the behavior of the individual. For example, blood pressure, disease complications, etc.
- In the case of internal time-dependent covariates, multi-state models are commonly used in the literature (Putter et al., 2007)

$$\lambda_{gh}(t|\mathbf{X}) = \lambda_{gh}(t)\exp(\beta_{gh}^{T}\mathbf{X}),$$

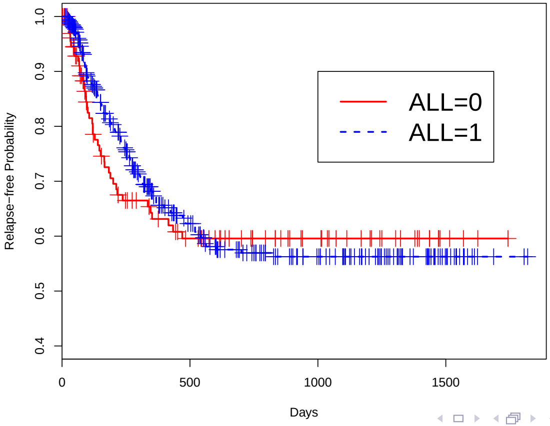
where $\lambda_{gh}(t|\mathbf{X})$ is the *transition hazard* from state g to h

Back to the EBMT Example



- Standard approach for analysis assumes proportional hazard models. This assumption implies that survival curves do not intersect
- In many applications, this assumption is not valid

EBMT Example: Empirical Kaplan-Meier Survival Curves of Relapse



Objectives

The objective of this work is

- to introduce a new general hazards model accommodating crossing hazards
- to account for internal time-dependent covariates using multi-state models

The model allows for the use of

1. non-constant hazards ratios
2. prediction of the probability of relapse

Proposed models (1): Time-independent Covariates

- General hazards model for the cause-specific hazards

$$\lambda_j(t|\mathbf{X}) = \lambda_j(t)\frac{\exp\{(\beta_j + \gamma_j)^T\mathbf{X}\}}{\exp(\beta_j^T\mathbf{X})F(t) + \exp(\gamma_j^T\mathbf{X})S(t)}, \quad j = 1, ..., k$$

  where β_j and γ_j are regression coefficients, $S(t) = \exp\left(-\sum_{j=1}^{k}\Lambda_j(t)\right)$ and $F(t) = 1 - S(t)$ are the baseline survival function and the baseline cumulative distribution function respectively, and $\Lambda_j(t)$ is the baseline cumulative hazard for the jth cause
- The general hazards model allows non-constant hazard ratios and has very appealing features

Proposed models (1): Time-independent Covariates(2)

It can be shown that for two sets of covariates $\mathbf{X}_1$ and $\mathbf{X}_2$

$$\frac{\lambda_j(t|\mathbf{X}_1)}{\lambda_j(t|\mathbf{X}_2)} \to \begin{cases} \exp\{\beta_j^T(\mathbf{X}_1 - \mathbf{X}_2)\}, & t \to 0 \\ \exp\{\gamma_j^T(\mathbf{X}_1 - \mathbf{X}_2)\}, & t \to \infty \end{cases}$$

- Therefore, β_j and γ_j can be interpreted as the short-term and long-term log-hazards ratios, respectively
- When $\beta_j = \gamma_j$, the general hazards model reduces to the Cox proportional hazards model

Likelihood

- Given n i.i.d. observations $\{(Y_i, D_i, \mathbf{X}_i), i = 1, ...n\}$, we can derive the following likelihood function for the unknown parameters $\theta \equiv \{(\beta_j, \gamma_j, \Lambda_j), j = 1, ..., K\}$

$$L_n(\theta) = \prod_{i=1}^{n} \prod_{j=1}^{K} \{\lambda(Y_i|\mathbf{X}_i)\}^{I(D_i=j)} \exp\{-\Lambda_j(Y_i|\mathbf{X}_i)\},$$

where

$$\Lambda_j(t|\mathbf{X}) = \int_0^t \frac{\exp\{(\beta_j + \gamma_j)^T \mathbf{X}\}}{\exp(\beta_j^T \mathbf{X})F(s) + \exp(\gamma_j^T \mathbf{X})S(s)} d\Lambda_j(s)$$

- To estimate the unknown parameters, we replace $\lambda_j(t)$ with the jump size of $\Lambda_j(\cdot)$ at time point t in the above likelihood function and then maximize the resultant nonparametric likelihood through the quasi-Newton algorithm

Proposed models (2): Internal Time-dependent Covariates

- In the case of internal time-dependent covariates, the multi-state model with transition-specific hazard becomes

$$\lambda_{gh}(t|\mathbf{X}) = \lambda_{gh}(t)\frac{\exp\{(\beta_{gh}+\gamma_{gh})^T\mathbf{X}\}}{\exp(\beta_{gh}^T\mathbf{X})F_g(t)+\exp(\gamma_{gh}^T\mathbf{X})S_g(t)}$$

  - where $\lambda_{gh}(t|\mathbf{X})$ is the conditional transition hazard from state g to state h given time-dependent covariates $\mathbf{X}$
  - and $S_g(t)$ is the baseline probability that the subject remains at state g at time t, and $F_g(t) = 1 - S_g(t)$
- The likelihood-based estimation procedure is applied for analyzing this model

Likelihood for the multi-state model

The likelihood function is given by

$$L_n(\theta) = \prod_{i=1}^{n} \prod_{gh \in \mathcal{C}} \{\lambda_{gh}(Y_{i,g}|\mathbf{X}_i)\}^{I(D_{i,gh}=1)} \exp\{-\Lambda_{gh}(Y_{i,g}|\mathbf{X}_i)\}$$

where $\mathcal{C}$ contains all possible transitions, and $D_{i,gh}$ indicates whether we observe the transition from state g to h for the ith individual

- This likelihood is maximized using quasi-Newton algorithm

Simulation study

We conducted simulation studies to examine the performance of the proposed methods under different scenarios

Simulations settings

Following are the steps to generate the data:

1. Generate a uniform random variable X in $[-1, 1]$
2. Generate a uniform random variable u in $[0, 1]$ and determine t such that $S(t|X) = u$, which is the generated failure time
3. Determine the cause of the failure.
   Generate a uniform random variable v in $[0, 1]$. Then

$$\text{cause} = \begin{cases} 1, & v \leq \frac{\lambda_1(t|X)}{\lambda_1(t|X)+\lambda_2(t|X)} \\ 2, & v > \frac{\lambda_1(t|X)}{\lambda_1(t|X)+\lambda_2(t|X)} \end{cases}$$

Simulations settings

We consider four scenarios for the values of regression parameters $\theta \equiv (\beta_1, \gamma_1, \beta_2, \gamma_2)$:

- (a) $(\beta_1, \gamma_1, \beta_2, \gamma_2) = (0.5, -0.5, -0.5, 0.5)$
- (b) $(\beta_1, \gamma_1, \beta_2, \gamma_2) = (0.5, -0, -0.5, 0)$
- (c) $(\beta_1, \gamma_1, \beta_2, \gamma_2) = (0, -0.5, -0, 0.5)$
- (d) $(\beta_1, \gamma_1, \beta_2, \gamma_2) = (0.5, 0.5, -0.5, -0.5)$

In all simulations, we set $\lambda_1(t) = 0.4$ and $\lambda_2(t) = 0.6$.
For each simulation scenario, we generated 1,000 replicates

Simulation: Results-Sample size= 300

Par	Bias	SE	SEE	CP	Bias	SE	SEE	CP
	$\theta = (0.5, -0.5, -0.5, 0.5)$				$\theta = (0.5, 0, -0.5, 0)$			
β_1	0.016	0.360	0.360	0.958	0.010	0.376	0.357	0.945
γ_1	-0.054	0.433	0.424	0.948	-0.017	0.439	0.427	0.949
β_2	-0.014	0.305	0.292	0.943	-0.010	0.300	0.288	0.955
γ_2	0.053	0.350	0.343	0.945	0.028	0.357	0.347	0.938
	$\theta = (0, -0.5, 0, 0.5)$				$\theta = (0.5, 0.5, -0.5, -0.5)$			
β_1	0.015	0.385	0.354	0.944	-0.007	0.361	0.358	0.957
γ_1	-0.046	0.474	0.448	0.947	0.033	0.474	0.464	0.949
β_2	-0.006	0.294	0.285	0.955	-0.003	0.297	0.289	0.956
γ_2	0.038	0.365	0.358	0.940	-0.006	0.387	0.371	0.947

SE, empirical standard deviation of the parameter estimates; SEE, average of standard error estimates; CP, coverage probability of the 95% confidence interval.

Data analysis

We applied the proposed methods to the EBMT example.

- Effects of ALL on the transition hazard from initial state to aGvHD, relapse or death

	short-term	long-term	Cox Model
Cause			
aGvHD	0.076(0.798)	-0.184(0.845)	0.019(0.91)
relapse	-6.167(0.021)	1.632(<0.001)	0.149(0.52)
death	-3.622(0.066)	1.192(0.055)	-0.381(0.14)

The p-values are included in the parentheses.

Data analysis (contd.)

- Effects of ALL on the transition hazard from the occurrence of aGvHD to relapse or death

	short-term	long-term	Cox Model
Cause			
relapse	-1.628(0.086)	0.401(0.512)	-0.553(0.06)
death	-0.094(0.883)	1.733(0.037)	0.521(0.07)

The p-values are included in the parentheses.

Data analysis (contd.)

- The proposed model detects significant short-term and long-term effects of ALL on the transition from the initial state to relapse which the Cox model fails to detect
- The proposed model detects significant long-term effect of ALL on the transition from aGvHD to death which the Cox model fails to detect

Summary

- Multi-state general hazards model in a competing risks setting
  - General hazards model allows departure from the proportional hazards assumptions and can model short-term and long-term covariate effects
  - Multi-state model is used to deal with internal time-dependent covariates
- Potential limitations of these models
  - The general hazards model may not work well if the time-varying effect is not monotone over time

Some references I

- Diao G, Zeng D. Semiparametric hazards rate model for modelling short-term and long-term effects. Submitted.
- Katsahian S, et al. The graft-versus-leukaemia effect after allogeneic bone-marrow transplantation: assessment through competing risks approaches. Statistics in medicine 2004; 24: 3851–63.